

# **As we may quote**

**Identifizieren und Zitieren in zukünftigen E-Science-  
Szenarien**

**Viele Fragen. Kaum Antworten ...**

Dr. Stefan Gradmann  
Universität Hamburg / Regionales Rechenzentrum  
[stefan.gradmann@rrz.uni-hamburg.de](mailto:stefan.gradmann@rrz.uni-hamburg.de)  
[www.rrz.uni-hamburg.de/RRZ/S.Gradmann](http://www.rrz.uni-hamburg.de/RRZ/S.Gradmann)





# Überblick

Funktionaler Kontext: German Academic Publishers (GAP),  
Hamburg UP, DINI

Im Zusammenhang mit dem Identifizieren und Zitieren von  
'Dokumenten' erwachsende Fragen zu Themenfeldern wie  
**Versionierung**, Dokumentintegrität, rights/restrictions  
management, politischer Glaubwürdigkeit, **Granularität**,  
**Szientometrie**

Was identifizieren wir eigentlich?

Ein semiologischer Exkurs

Abschließende Fragen zum konzeptionellen Instrumentarium

Object Creation: GAP, Hamburg UP. Electronic Open Access publishing of scientific 'articles' **and** 'book-like' objects (and thus in a STM **and** SSH perspective!)

([www.gap-portal.de](http://www.gap-portal.de) and [www.hup.rrz.uni-hamburg.de](http://www.hup.rrz.uni-hamburg.de))

Object Repositories: DINI Repository Building, Repository Federation, Interoperability of Repositories and Repository Certification.

DINI Certificate <http://www.dini.de/documents/Zertifikat-en.pdf>

„A persistent identifier must be assigned to every document on the repository. A possible solution is the assignment of URNs.“

„Persistent identifiers (PI) can additionally be used to support *harvesting*. PIs can be used as *workflow-control mechanisms*, where unambiguous identification of objects is required, as in e. g. *de-centralized processes*, *checking for doubles*, *authentication mechanisms*, *alarm systems*, or for *multimedia or modular objects*. *Objects* whose modules are stored in different repositories *can be merged* effectively through the use of PIs. By selecting clearly identified modules, different *versions can be generated* without having to store each of these versions. Digital objects can be *cited* through their PIs in digital environments as well as in non-digital environments.“

DINI certificate statement: „Every alteration of a document creates a new document. A persistent identifier is assigned to this new document. Versioning allows for identification of previous versions of a document. Metadata work as links between generations of a document.“

Was genau ist in einem solchen Szenario 'persistent': das **Dokument / Objekt**? Der **Identifier**? Der **Link** zwischen beiden?  
=> Generalfrage: Was identifiziert wir eigentlich?



# Integrity & Authentication

Integrity of instances: Identification and Authentication of information objects and actors.

DINI certificate: „To guarantee the integrity of a document for the reader, a hash must exist for every document in the repository. Changing demands on security foster technical developments in cryptography; algorithms for hash numbers have to be adapted according to these new developments. Current recommendations include the Message Digest Algorithm MD5 or the Secure Hash Algorithm SHA-1. Application of cryptography algorithms should be publicized for the readers' information.“

Who is entitled to alter document integrity statements?

How is this person/instance authenticated?

How is the scope of his/its authorization determined?

# Rights/Restrictions Management

What is protected by rights / restrictions management (DRM):  
objects alone or objects together with access methods?

Do we consider frameworks such as



or



-21

as guiding lines, and to what extent do such choices affect  
persistency of the overall framework?

And can identifiers and DRM technology framework be separated  
effectively and neatly over time?

If this separation cannot be guaranteed, the consequences in  
terms of persistency might well be catastrophic, especially in  
highly interwoven scientific linking scenarios!

# Social / Political Dimension



Stu Weibel  
Cork 2004:



„Technology can help, but the problem is, at its heart, a social one“ (Stu Weibel, 2004)

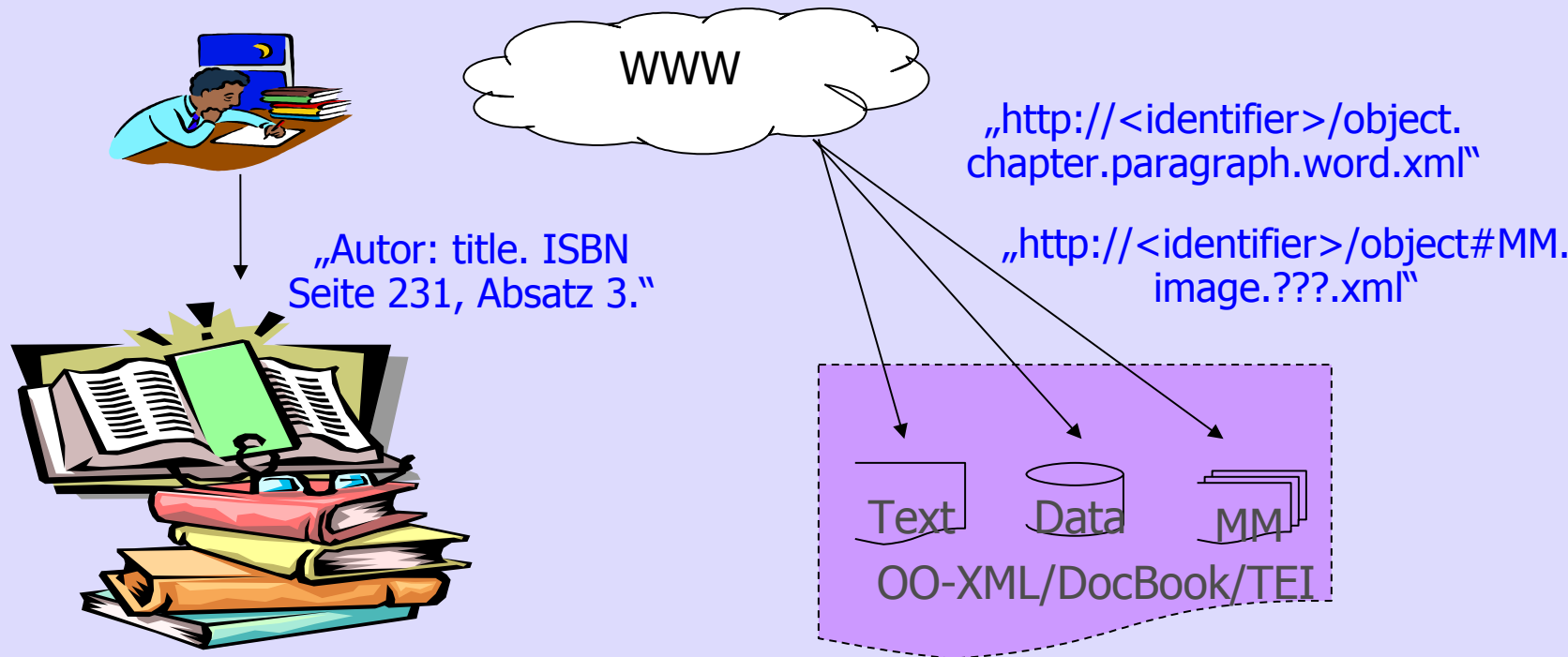
„At the end of the day, the only guarantee of the usefulness and persistence of identifier systems is the commitment of the organizations which assign, manage, and resolve identifiers.“

Who do we trust? And who should apply identifiers as a consequence (authors, institutions, companies)?

Prevailing choice within DINI currently seems to be URN:NBN, with German National Library (DDB) as hosting institution.

But is that choice sufficient? Think about objects that are not within DDBs „bibliographic“ scope.

# Granularität und Zitieren / Referenzieren



Wie granular muß der Identifier sein, um wissenschaftliche Zitate zu ermöglichen und sie langfristig haltbar zu machen?

Kann diese Granularität variabel implementiert werden, so daß sie abhängig von der Komplexität der referenzierten Objekte skaliert? Absolutes vs. relatives Referenzieren!

Identifizieren: **work**, **expression** oder **manifestation**?

Welche Granularität ist für heutige und zukünftig denkbare szientometrische Verfahren erforderlich?



# Was identifizieren wir eigentlich?

Welche Entitäten identifizieren wir eigentlich?

Technisch gesprochen: Dateien? Bits & bytes?

Dokumente? Daten? Lokationen? Text? Bilder?

Semantisch gesprochen: Konzepte?

Informationen? Zeichen?

Und welcher Art ist die Verbindung zwischen

Identifizier und Objekt: Link? Signifikanz?

Beschreibung? Repräsentation? Surrogat?

Und dabei bleiben echte epistemologische

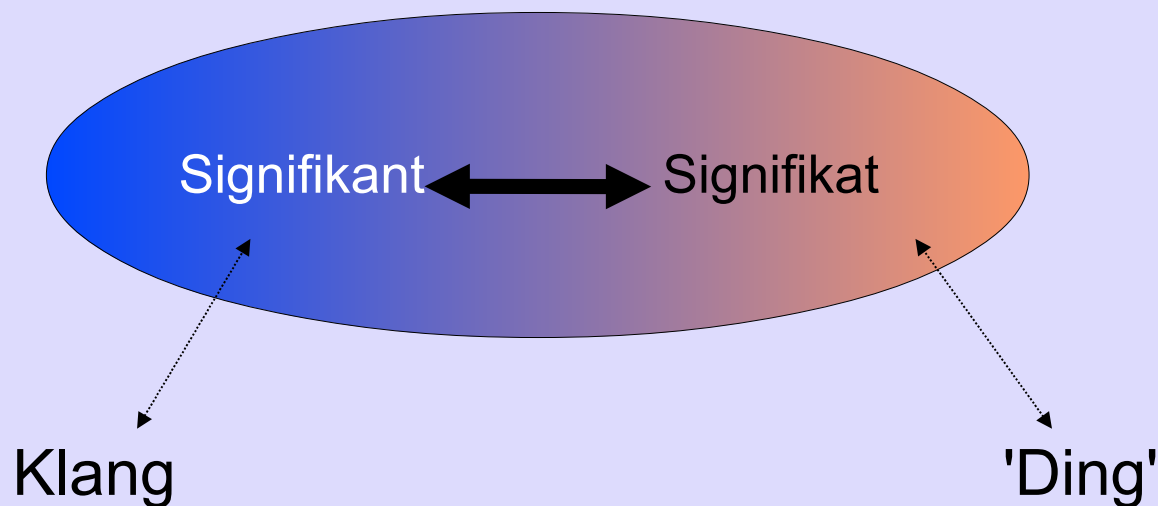
Sprengsätze wie die Frage nach dem Konzept  
'Identität' selbst noch ausgeblendet ...

'Pointer -> Objekt'-Modell



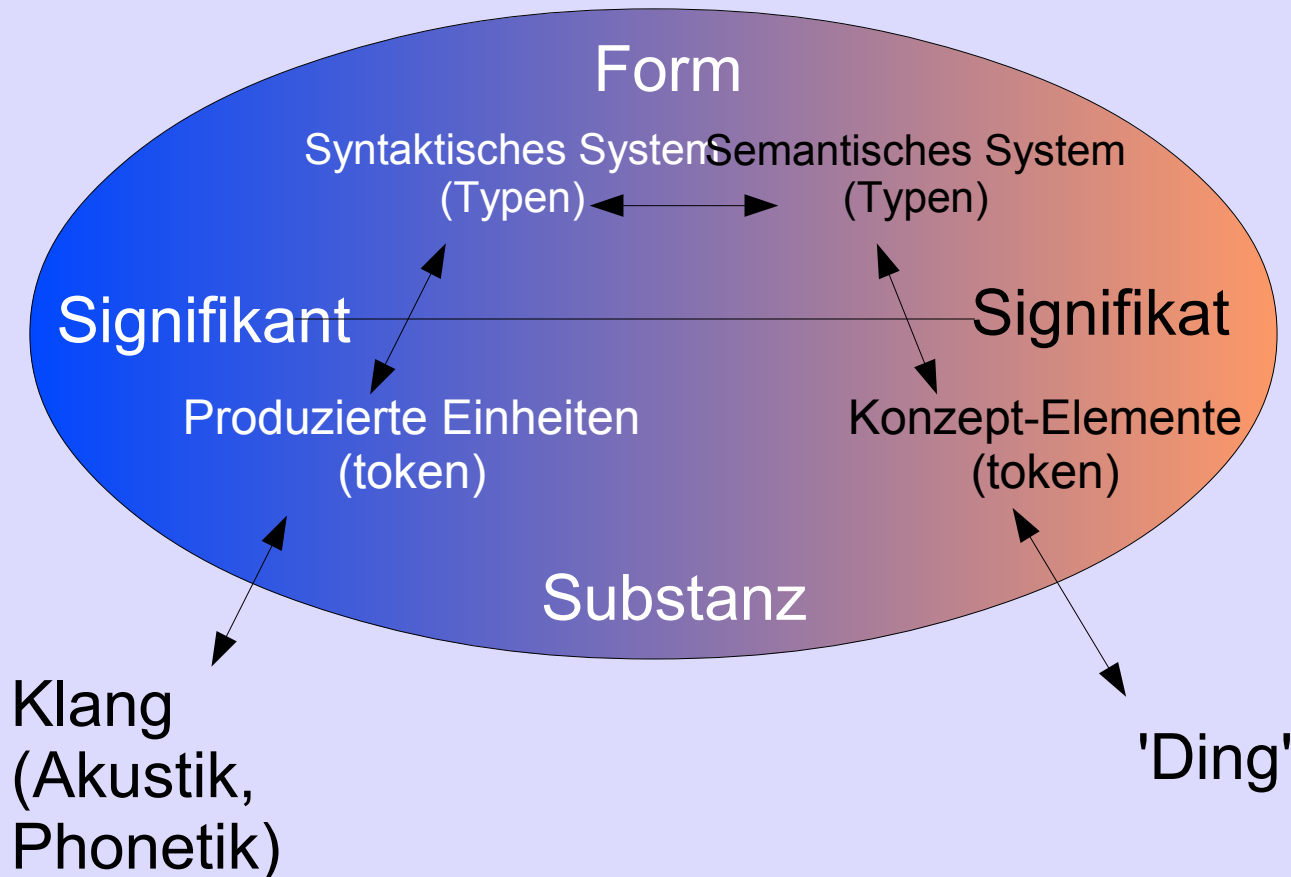
Aber: /Messer/, /couteau/, /knife/?

Von daher Saussure's Modell des 'Sprachzeichens':



# Semiologischer Exkurs / 2

Mit Erweiterungen von Hjelmslev, Eco und anderen:





# Noch ein paar Fragen zum Schluss ...

Der semiologische Exkurs ist nicht als Vorschlag für ein neues methodologisches Instrumentarium zu verstehen. Er ermöglicht es jedoch, eine Reihe schon angedeuteter Fragen präziser zu stellen:

Konzipieren wir 'identifizier' als '**Zeiger**', '**locators**', als '**Namen**', als '**referentielle Surrogate**' oder als '**Zeichen**'?

Sollten wir sie als '**Zeiger**', als '**label**' oder als '**Metadaten-Attribute**' begreifen?

Sollten Identifizier eine '**Bedeutung**' haben?

Wenn **ja**: leiten wir seine Semantik aus dem referenzierten **Objekt** oder aus dem **Verwendungskontext** ab ?

Wenn **nein**: Wie stellen wir die Dauerhaftigkeit der Verbindung zwischen **Identifizier** und **Objekten** und derjenigen zwischen **Identifiers** und **Verwendungskontext** sicher?

Solange diese und verwandte Fragen nicht beantwortbar werden kann genuin digital basiertes wissenschaftliches Zitieren nicht funktionieren! Und 'E-Science'-Umgebungen bleiben dann –  
12 zumindest vorläufig – rein experimentelle Spielumgebungen!!

# **Elektronisches Publizieren & Open Access in den 'Geisteswissenschaften'**

**Gedanken. Splitter.**

Dr. Stefan Gradmann  
Universität Hamburg / Regionales Rechenzentrum  
[stefan.gradmann@rrz.uni-hamburg.de](mailto:stefan.gradmann@rrz.uni-hamburg.de)  
[www.rrz.uni-hamburg.de/RRZ/S.Gradmann](http://www.rrz.uni-hamburg.de/RRZ/S.Gradmann)





# Ungeordnete Gedanken

Einige Worte zu den 'zwei Kulturen' ...

Wie können 'Elektronisches Publizieren' und 'Open Access' in den hermeneutisch geprägten Disziplinen funktionieren?

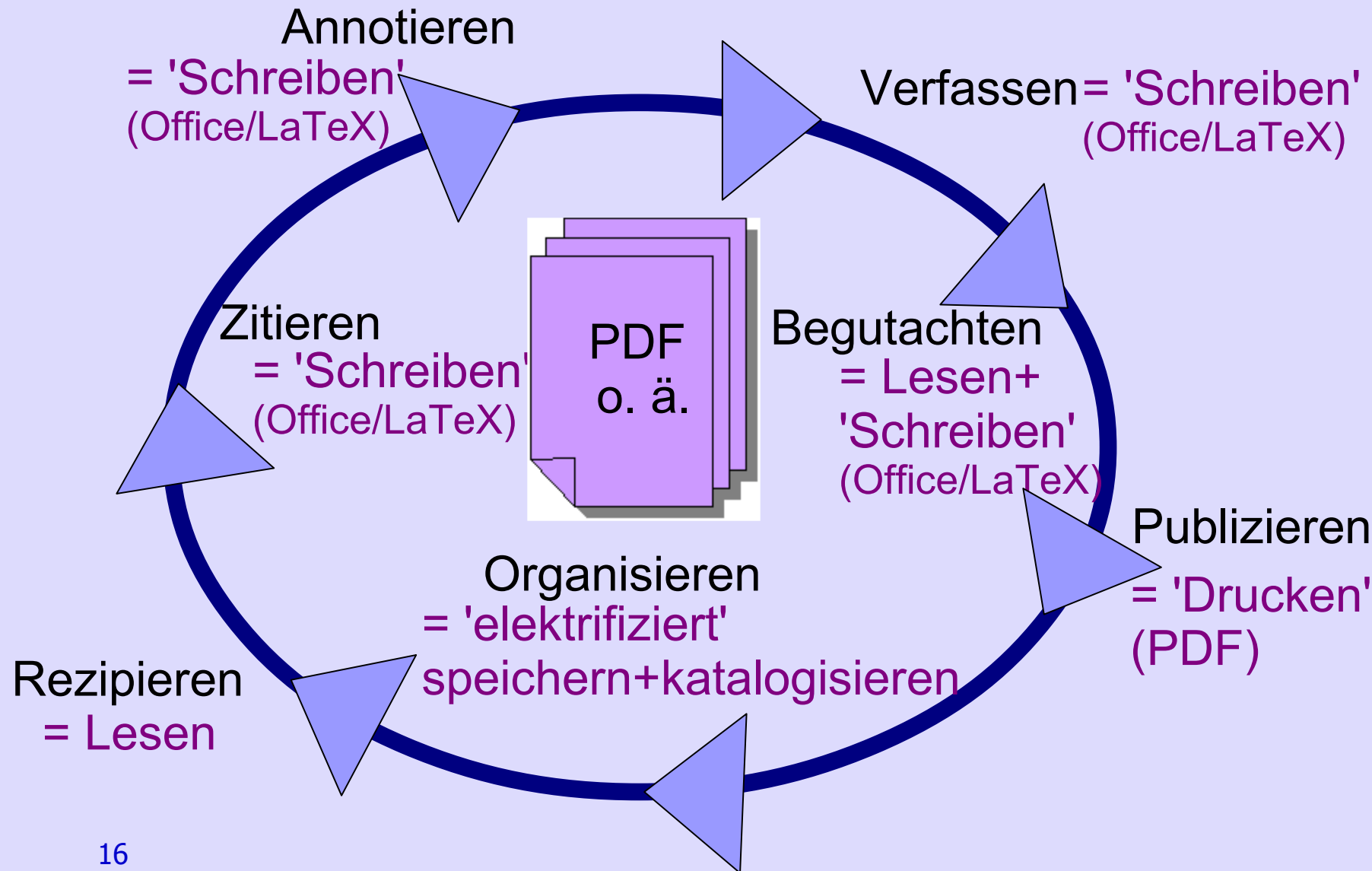
Was verändern 'Elektronisches Publizieren' und 'Open Access' in den hermeneutisch geprägten Disziplinen?

Um diese beiden Fragen zu beantworten ist ein Verständnis für die spezifische Verbindung von 'Form' und 'Inhalt' in den hermeneutisch geprägten Wissenschaften erforderlich!

# Informationskontinuum in traditioneller Funktionsweise

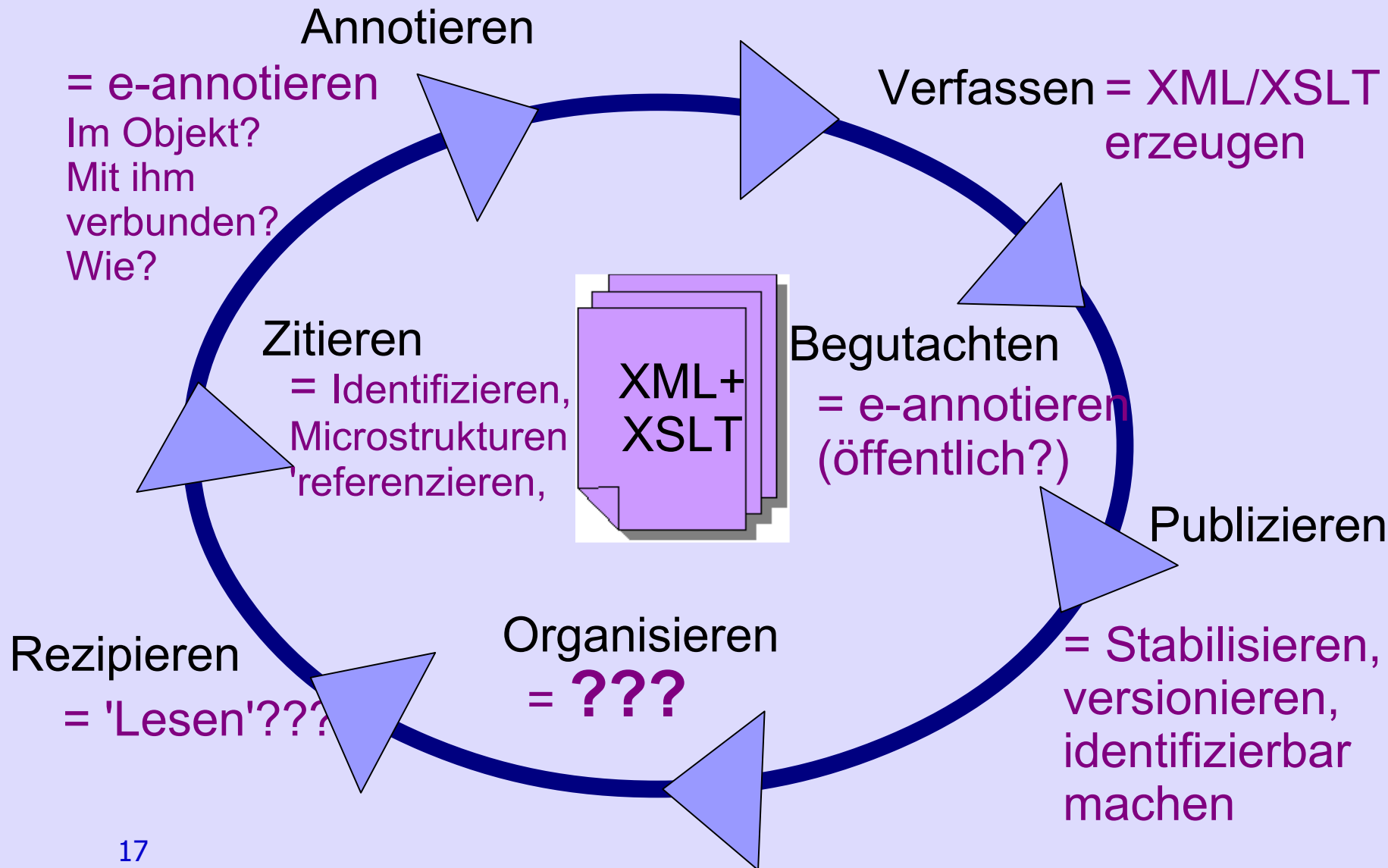


# Informationskontinuum im Emulationsmodus





# Informationskontinuum: going digital



# Bedeutungsmodi in der Wissenschaftskommunikation

Bedeutungsmodi und Dokumentmodellierung im E-Science-Kontext sind bislang vorrangig geprägt durch das in den empirischen Wissenschaften dominante Informationsmodell:

**Forschung => 'Resultate' => 'Packaging' => Publikation**

Ein robustes und wenig komplexes 'Container'-Modell

Digitale Wissenschaftskommunikation  $\approx$  Elektrifizierte Wissenschaftskommunikation

Informationsmodellierung in den hermeneutisch geprägten Geistes- und Sozialwissenschaften findet in einem substantiell anderen Bedeutungsmodus statt:

**(Forschung  $\Leftrightarrow$  diskursives 'Packaging') => Publikation**

Komplexe Dokumentmodelle, die mit den wissenschaftlichen Kernprozessen eng verwoben sind

Komplexe Zeichenbeziehungen sind konstitutiv.

'Container'-Modelle sind reduktionistisch und unangemessen.

# Konsequenzen für Elektronisches Publizieren und OA

Open Access verändert in den Geisteswissenschaften prima facie sehr wenig (=> Ducasse, De Gruyter, Harnad)

Der Umgang mit nicht-print-analogen, nicht-linearen, genuin digitalen Informationsobjekten (und damit auch das Elektronische Publizieren) tangieren das Informationskontinuum aller Wissenschaften – würden jedoch dasjenige der hermeneutisch geprägten Disziplinen in besonderer Weise verändern (=> McCarty)

=> E-Science wird erst wirklich ernstzunehmen sein, wenn sie die Herausforderung der Implementierung von 'Verstehen' wenigstens ein Stückweit gemeistert hat (und der Unsworth-Report greift deutlich zu kurz!)

**VS.**

=> E-Science wird ohne Rücksicht auf komplexe, hermeneutisch geprägte Bedeutungsmodi 'marschieren' (und diese allenfalls 'gekapselt' mitnehmen – und der Unsworth-Report wäre der adäquate Versuch, auf diesen Zug zu springen)

# Ein Nicht-Lineares Informationskontinuum

